



PATENT ABSTRACTS OF JAPAN

(11) Publication number: 2001013986 A

(43) Date of publication of application: 19.01.2001

(51) Int. Cl. G10L 15/10
G10L 15/06, G10L 15/14

(21) Application number: 11182650

(22) Date of filing: 29.06.1999

(71) Applicant: NEC CORP

(72) Inventor: ISHIKAWA SHINYA

(54) **SPEAKER ADAPTIVE DEVICE AND METHOD OF VOICE RECOGNITION ACOUSTIC MODEL**

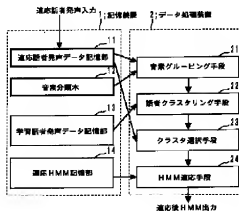
(57) Abstract:

PROBLEM TO BE SOLVED: To obtain a device, in which finer speaker adaptive is made possible, by conducting speaker clustering for every fine group of phoneme, generating and selecting an adequate speaker cluster.

SOLUTION: The device is provided with a phoneme grouping means 21, which finely conduct phoneme grouping in accordance with data, a speaker clustering means 22, which conducts individual clustering and selecting for each phoneme group, and a cluster selecting means 23 and speaker adaptive is realized by speaker cluster selection for individual phoneme. Note that speaker clustering is conducted for every fine phoneme group in accordance with the uttering of an

adaptive speaker and a cluster is selected. As a result, learning speaker uttering data, that are more closer to an adaptive speaker, is finely collected.

COPYRIGHT: (C)2001,JPO



(51) Int.Cl. ⁷	識別記号	F I	テマコード [*] (参考)
G 1 0 L 15/10		G 1 0 L 3/00	5 3 1 K 5 D 0 1 5
15/06			5 2 1 S
15/14			5 2 1 V
			5 3 5 Z

審査請求 有 請求項の数11 O L (全 8 頁)

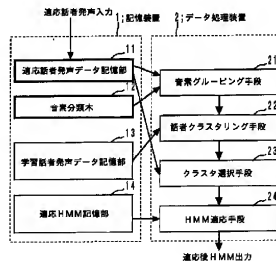
(21) 出願番号	特願平11-182650	(71) 出願人	000004237 日本電気株式会社 東京都港区芝五丁目7番1号
(22) 出願日	平成11年6月29日 (1999.6.29)	(72) 発明者	石川 晋也 東京都港区芝五丁目7番1号 日本電気株式会社内
		(74) 代理人	100080816 弁理士 加藤 朝道 Fターム(参考) 5D015 AA03 BB02 GG01 GG03 HH06 HH23

(54) 【発明の名称】 音声認識用音素モデルの話者適応装置及び方法

(57) 【要約】

【課題】 音素のできるだけ細かいグループ毎に話者クラスタリングを行ない、適切な話者クラスタを作成し選択することで、より緻密な話者適応を可能とする装置の提供。

【解決手段】 音素のグループ分けをデータに応じて細かく行なう音素グループピング手段 (図1の21) と、各音素グループに対し別個にクラスタリング、選択を行う話者クラスタリング手段 (図1の22) と、クラスタ選択手段 (図1の23) とを備え、音素別の話者クラスタ選択による話者適応を実現する。



【特許請求の範囲】

【請求項1】新話者の音声（「適応話者発声」という）を用いて話者用の標準パターンを自動作成する話者適応化装置において、

適応話者の適応発声データを音素のグループに分ける音楽グルーピング手段と、

各音楽グループ毎に複数の標準パターンを記憶する記憶手段と、

前記音楽グループに対応する適応話者発声との類似性から標準パターンを選択する選択手段と、
を備えたことを特徴とする話者適応装置。

【請求項2】新話者の音声（以下「適応話者発声」）を用いてその話者用の標準パターンを自動作成する話者適応化装置において、

適応話者の適応発声データを音素のグループに分ける音楽グルーピング手段と、

各音楽グループ毎に複数の学習発声情報を記憶する記憶手段と、

前記音楽グループに対応する適応話者発声との類似性から学習発声情報を選択する選択手段と、

を備えたことを特徴とする話者適応装置。

【請求項3】請求項1又は2に記載の話者適応装置において、音楽グループ毎の複数の標準パターンまたは学習発声情報を、話者クラスタリングによって作成する手段を備えたことを特徴とする話者適応装置。

【請求項4】前記音楽グルーピング手段が、音楽のグループ分けにおいて、音楽の階層的な分類を用いる、ことを特徴とする請求項1乃至3のいずれかに記載の話者適応装置。

【請求項5】前記音楽グルーピング手段が、音楽のグループ分けにおいて、各音楽グループにおける適応話者の発声量に応じて音楽分類木上で適切な階層のグループ分けを選ぶ、ことを特徴とする請求項1乃至3のいずれかに記載の話者適応装置。

【請求項6】音声入力手段から入力されデジタル信号に変換されHMM (Hidden Markov Model: 隠れマルコフモデル) を適応する対象の適応話者の発声データを記憶する適応話者発声データ記憶部と、

音楽の分類情報が予め記憶されている音楽分類木記憶部と、

複数話者の発声が予め記憶されている学習話者発声データ記憶部と、

適応する元のHMMが記憶されている適応HMM記憶部と、

適応話者発声データを前記適応話者発声データ記憶部より読み出し、発声数の少ない音楽について、類似している音楽を、前記音楽分類木記憶部の分類に基づきグループ化し、複数の音楽グループを出力する音楽グルーピング手段と、

前記音楽グルーピング手段から出力された複数の音楽グ

ループのそれぞれについて、学習話者発声データを前記学習話者発声データ記憶部より読み出して、話者単位で、クラスタリングを行ない、クラスタ化された学習話者発声データを、音楽グループ毎に出力する話者クラスタリング手段と、

音楽グループ毎に、該音楽グループに対応する適応話者発声データを、前記適応話者発声データ記憶部より読み出して、前記話者クラスタリング手段から出力された学習話者発声データのクラスタのそれぞれと比較し、最適なクラスタを出力するクラスタ選択手段と、
音楽グループ毎の学習話者発声データクラスタを用いて、適応元のHMMを、前記適応HMM記憶部より読み出し、適応を行ないHMMを出力するHMM適応手段と、

を備えたことを特徴とする話者適応装置。

【請求項7】音声入力手段から入力されデジタル信号に変換されHMM (Hidden Markov Model: 隠れマルコフモデル) を適応する対象の適応話者の発声データを記憶する適応話者発声データ記憶部と、

音楽の分類情報、及び、話者クラスタリング情報を記憶する音楽分類木及びクラスタ情報記憶部と、

複数話者の発声が予め記憶されている学習話者発声データ記憶部と、

適応する元のHMMが記憶されている適応HMM記憶部と、

入力された適応話者発声データを前記適応話者発声データ記憶部より得て、発声数の少ない音楽について類似している音楽を前記音楽分類木及びクラスタ情報記憶部の分類に基づきグループ化し、複数の音楽グループと同時に、前記音楽分類木及びクラスタ情報記憶部から得た前記音楽グループに対する話者クラスタリング情報を出力する音楽グルーピング選択手段と、

前記音楽グルーピング選択手段から出力された音楽グループ及び話者クラスタリング情報毎に、前記音楽グループに対応する適応話者発声データを前記適応話者発声データ記憶部より読み出し、前記話者クラスタリング情報に対して必要に応じて学習話者発声データを前記学習話者発声データ記憶部から読み出し、クラスタごとにHMMを構成し、最適なクラスタの情報を出力するクラスタ選択手段と、

前記クラスタ選択手段から出力された音楽グループ毎の学習話者発声データクラスタの情報から、必要に応じて学習データを前記学習話者発声データ記憶部から読み出し、適応元のHMMを前記適応HMM記憶部より読み出し、適応を行なってHMMを出力するHMM適応手段と、

を備えたことを特徴とする話者適応装置。

【請求項8】(a) 適応話者発声データを、HMM (Hidden Markov Model: 隠れマルコフモデル) を適応する対象の適応話者の発声データを記憶する適応話者発声

データ記憶部より読み出し、発声数の少ない音素について、類似している音素を、音素の分類情報が予め記憶されている音素分類木記憶部の分類に基づきグループ化し、複数の音素グループを出力するステップと、

(b) 前記ステップ (a) で出力された複数の音素グループのそれぞれについて、学習話者発声データを、複数話者の発声が予め記憶されている学習話者発声データ記憶部より読み出して、話者単位で、クラスタリングを行ない、クラスタ化された学習話者発声データを、音素グループ毎に出力するステップと、

(c) 音素グループ毎に、該音素グループに対応する適応話者発声データを、前記適応話者発声データ記憶部より読み出して、入力された学習話者発声データのクラスタのそれぞれと比較し、最適なクラスタを出力するステップと、

(d) 音素グループ毎の学習話者発声データクラスタを用いて、適応元のHMMを、適応する元のHMMが記憶されている適応HMM記憶部より読み出し、適応を行ないHMMを出力するステップと、

を含むことを特徴とする話者適応方法。
【請求項9】 (a) 入力された適応話者発声データを、HMM (Hidden Markov Model: 隠れマルコフモデル) を適応する対象の適応話者の発声データを記憶する適応話者発声データ記憶部より得て、発声数の少ない音素について似ている音素を、音素の分類情報、及び話者クラスタリング情報を記憶する音素分類木及びクラスタ情報記憶部の分類に基づきグループ化し、複数の音素グループと同時に、前記音素分類木及びクラスタ情報記憶部から得た音素グループに対する話者クラスタリング情報を出力するステップと、

(b) ステップ (a) で出力された音素グループ及び話者クラスタリング情報毎に、前記音素グループに対応する適応話者発声データを、複数話者の発声が予め記憶されている適応話者発声データ記憶部より読み出し、話者クラスタリング情報に対して必要に応じて学習話者発声データを前記学習話者発声データ記憶部から読み出し、クラスタごとにHMMを構成し、最適なクラスタの情報を出力するステップと、

(c) 入力の音素グループ毎の学習話者発声データクラスタの情報から、必要に応じて学習データを前記学習話者発声データ記憶部から読み出し、適応元のHMMを、適応する元のHMMが記憶されている適応HMM記憶部より読み出し、適応を行ないその結果のHMMを出力するステップと、

を含むことを特徴とする話者適応方法。

【請求項10】 音声入力手段から入力されデジタル信号に変換されHMM (Hidden Markov Model: 隠れマルコフモデル) を適応する対象の適応話者の発声データを記憶する適応話者発声データ記憶部と、音素の分類情報が予め記憶されている音素分類木記憶部

と、複数話者の発声が予め記憶されている学習話者発声データ記憶部と、

適応する元のHMMが記憶されている適応HMM記憶部と、

データ処理装置と、を備えた話者適応装置において、

(a) 適応話者発声データを前記適応話者発声データ記憶部より読み出し、発声数の少ない音素について、類似している音素を、前記音素分類木記憶部の分類に基づきグループ化し、複数の音素グループを出力する音素グループ化処理と、

(b) 前記音素グループ化処理から出力された複数の音素グループのそれぞれについて、学習話者発声データを前記学習話者発声データ記憶部より読み出して、話者単位で、クラスタリングを行ない、クラスタ化された学習話者発声データを、音素グループ毎に出力する話者クラスタリング処理と、

(c) 音素グループ毎に、その音素グループに対応する適応話者発声データを、前記適応話者発声データ記憶部より読み出して、入力された学習話者発声データのクラスタのそれぞれと比較し、最も適当なクラスタを出力するクラスタ選択処理と、

(d) 音素グループ毎の学習話者発声データクラスタを用いて、適応元のHMMを、前記適応HMM記憶部より読み出し、適応を行ないHMMを出力するHMM適応処理と、

の前記 (a) 乃至 (d) の処理を前記データ処理装置で実行させるためのプログラムを記録した記録媒体。

【請求項11】 音声入力手段から入力されデジタル信号に変換されHMM (Hidden Markov Model: 隠れマルコフモデル) を適応する対象の適応話者の発声データを記憶する適応話者発声データ記憶部と、音素の分類情報、及び、話者クラスタリング情報を記憶する音素分類木及びクラスタ情報記憶部と、複数話者の発声が予め記憶されている学習話者発声データ記憶部と、

適応する元のHMMが記憶されている適応HMM記憶部と、

データ処理装置とを備えた話者適応装置において、

(a) 入力された適応話者発声データを前記適応話者発声データ記憶部より得て、発声数の少ない音素について似ている音素を前記音素分類木及びクラスタ情報記憶部の分類に基づきグループ化し、複数の音素グループと同時に、前記音素分類木及びクラスタ情報記憶部から得たそれに対する話者クラスタリング情報を出力する音素グループ化選択処理と、

(b) 前記音素グループ化選択処理より受け取った音素グループ及び話者クラスタリング情報毎に、前記音素グループに対応する適応話者発声データを前記適応話者発声データ記憶部より読み出し、話者クラスタリング情

報に対して必要に応じて学習話者発声データを前記学習話者発声データ記憶部から読み出し、クラスごとにHMMを構成し、もっとも適当なクラスの情報を出力するクラス選択処理と、

(c) 前記クラス選択処理で出力された音素グループ毎学習話者発声データクラスの情報から、必要に応じて学習データを前記学習話者発声データ記憶部から読み出し、適応元のHMMを前記適応HMM記憶部より読み出し、適応を行ないその結果のHMMを出力するHMM適応処理と、

の前記(a)乃至(d)の処理を前記データ処理装置で実行させるためのプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、音声認識用音素モデルを用いた話者適応装置に関し、特に適応発声に応じた話者クラスタリングデータを用いた話者適応装置及び方法に関する。

【0002】

【従来の技術】 話者クラスタリングを用いた話者適応手法、音声認識における音素モデルを特定の話者に適応させる手法として、例えば「電子情報通信学会論文誌 Vol. J78-D-II No. 1の第1頁〜第9頁、1995年」に掲載された「木構造話者クラスタリングを用いた話者適応」と題する論文が参照される。

【0003】 この種の話者クラスタリングは、上記論文にも記載されているように、あらかじめ多数の話者について各話者の発声で作成された特定話者用標準パターンを作成し、標準パターン間の距離を定義した上、K-mean法など公知のクラスタリング手法を用いて、話者単位のクラスタリングを行なっている。

【0004】 また話者クラスタリングを用いた話者適応は、クラス毎に属するすべての話者発声を用いて学習された標準パターンを、適応話者発声を用いて尤度などで選択し、これを適応後の標準パターンとしている。

【0005】

【発明が解決しようとする課題】 しかしながら、上記した従来の手法は下記記載の問題点を有している。

【0006】 話者のすべての発声(音素)を一まとめにして分類している、ということである。

【0007】 これに対し、ある話者Aと話者Bは音素1の発声に対しては似ているが、音素2の発声に関しては話者Aは話者Bとは似ていない、別の話者Cと似ている、という場合もあり得る。

【0008】 このように、音素毎に話者クラスタリングの最適なクラス分けが異なる可能性がある。このため、クラスタリング結果は、おどろばになり、良い話者適応ができない可能性がある。

【0009】 したがって本発明は、上記問題点に鑑みてなされたものであって、その目的は、音素のできるだけ

細かいグループ毎に話者クラスタリングを行ない、適切な話者クラスを作成し選択することで、より精密な話者適応を可能とする装置及び方法を提供することにある。

【0010】

【課題を解決するための手段】 前記目的を達成する本発明の新話者の音声(「適応話者発声」という)を用いて話者用の標準パターンを自動作成する話者適応化装置において、適応話者の適応発声データを音素のグループに分ける音素グループビンギング手段と、各音素グループ毎に、複数の標準パターンを記憶する記憶手段と、前記音素グループに対応する適応話者発声との類似性から標準パターンを選択する選択手段と、を備える。

【0011】 また、本発明は、適応話者の適応発声データを音素のグループに分ける音素グループビンギング手段と、各音素グループ毎に、複数の学習発声情報を記憶する記憶手段と、前記音素グループに対応する適応話者発声との類似性から学習発声情報を選択する選択手段と、を備える。

【0012】 本発明においては、音素グループ毎の複数の標準パターンまたは学習発声情報を、話者クラスタリングによって作成する。

【0013】 また、本発明において、前記音素グループビンギング手段が、音素のグループ分けにおいて、音素の階層的な分層を用いる。あるいは、前記音素グループビンギング手段が、音素のグループ分けにおいて、各音素グループにおける適応話者の発声基に応じて音素分層本上で適切な階層的なグループ分けを選ぶ。

【0014】

【発明の実施の形態】 本発明の実施の形態について説明する。本発明は、話者の発声を一まとめにして話者クラスタリングするのではなく、音素のグループ毎に、話者クラスタリングを行ない、音素の適切な話者クラスを作成して選択することで、より精密な話者適応を行なうようにしたものである。本発明の話者適応装置は、その好ましい実施の形態において、音素のグループ分けを適応発声データに応じて細かく行なう音素グループビンギング手段(図1の21)と、各音素グループに対して個別にクラスタリング、選択を行う話者クラスタリング手段(図1の22)と、クラス選択手段(図1の23)とを備え、音素別の話者クラス選択による話者適応を実現している。

【0015】 本発明の話者適応装置は、すべての音素のグループ分けに対して、予め話者クラスタリングを行って記憶装置の記憶しておくようにしてもよい。

【0016】 本発明は、その好ましい実施の形態において、HMM(Hidden Markov Model: 隠れマルコフモデル)を適応する対象の適応話者の発声データを記憶する適応話者発声データ記憶部と、音素の分類情報が予め記憶されている音素分類木記憶部と、複数の話者の発声が

予め記憶されている学習話者発声データ記憶部と、適応する元のHMMが記憶されている適応HMM記憶部と、適応話者発声データの前記適応話者発声データ記憶部より読み出し、発声数の少ない音素について、類似している音素を、前記音素分類木記憶部の分類に基づきグループ化し、複数の音素グループを出力する音素グルーピング手段(21)と、前記音素グルーピング手段から出力された複数の音素グループのそれぞれについて、学習話者発声データを前記学習話者発声データ記憶部より読み出して、話者単位で、クラスタリングを行ない、クラスタ化された学習話者発声データを、音素グループ毎に出力する話者クラスタリング手段(22)と、音素グループ毎に、その音素グループに対応する適応話者発声データを、前記適応話者発声データ記憶部より読み出して、入力された学習話者発声データのクラスタのそれぞれと比較し、最も適当なクラスタを出力するクラスタ選択手段(23)と、音素グループ毎の学習話者発声データクラスタを用いて、適応元のHMMを、前記適応HMM記憶部より読み出し、適応を行ないHMMを出力するHMM適応手段(24)と、を備える。これらの各手段21～24は、データ処理装置で実行されるプログラムによりその処理が実現される。この場合、該プログラムを記録した記録媒体もしくは通信媒体から該プログラムを読み出し、データ処理装置で実行すること本発明を実施することができる。

【0017】また本発明は、その好ましい実施の形態において、HMM (Hidden Markov Model: 隠れマルコフモデル) を適応する対象の適応話者の発声データを記憶する適応話者発声データ記憶部と、音素の分類情報及び、話者クラスタリング情報を記憶する音素分類木及びクラスタ情報記憶部と、複数話者の発声が予め記憶されている学習話者発声データ記憶部と、適応する元のHMMが記憶されている適応HMM記憶部と、入力された適応話者発声データを前記適応話者発声データ記憶部より得て、発声数の少ない音素について似ている音素を前記音素分類木及びクラスタ情報記憶部の分類に基づきグループ化し、複数の音素グループと同時に、前記音素分類木及びクラスタ情報記憶部から得たそれに対する話者クラスタリング情報を出力する音素グルーピング選択手段(21A)と、受け取った音素グループ及び話者クラスタリング情報毎に、前記音素グループに対応する適応話者発声データを前記適応話者発声データ記憶部より読み出し、話者クラスタリング情報に対して必要に応じて学習話者発声データを前記学習話者発声データ記憶部から読み出し、クラスタごとにHMMを構成し、もっとも適当なクラスタの情報を出力するクラスタ選択手段(23)と、入力の音素グループ毎学習話者発声データクラスタの情報から、必要に応じて学習データを前記学習話者発声データ記憶部から読み出し、適応元のHMMを前記適応HMM記憶部より読み出し、適応を行ないその結

果のHMMを出力するHMM適応手段(24)と、を備える。これらの各手段21A、23、24は、データ処理装置で実行されるプログラムによりその処理が実現される。この場合、該プログラムを記録した記録媒体もしくは通信媒体から該プログラムを読み出し、データ処理装置で実行すること本発明を実施することができる。

【0018】また本発明の方法は、以下のステップよりなる。

【0019】ステップ1：適応話者発声データを、HMM (Hidden Markov Model: 隠れマルコフモデル) を適応する対象の適応話者の発声データを記憶する適応話者発声データ記憶部より読み出し、発声数の少ない音素について、類似している音素を、音素の分類情報に基づき記憶されている音素分類木記憶部の分類に基づきグループ化し、複数の音素グループを出力する。

【0020】ステップ2：前記ステップ1で出力された複数の音素グループのそれぞれについて、学習話者発声データを、複数話者の発声が予め記憶されている学習話者発声データ記憶部より読み出して、話者単位で、クラスタリングを行ない、クラスタ化された学習話者発声データを、音素グループ毎に出力する。

【0021】ステップ3：音素グループ毎に、その音素グループに対応する適応話者発声データを、前記適応話者発声データ記憶部より読み出して、入力された学習話者発声データのクラスタのそれぞれと比較し、最も適当なクラスタを出力する。

【0022】ステップ4：音素グループ毎の学習話者発声データクラスタを用いて、適応元のHMMを、適応する元のHMMが記憶されている適応HMM記憶部より読み出し、適応を行ないHMMを出力する。

【0023】また本発明の方法は、以下のステップよりなる。

【0024】ステップ1：入力された適応話者発声データを、HMM (Hidden Markov Model: 隠れマルコフモデル) を適応する対象の適応話者の発声データを記憶する適応話者発声データ記憶部より得て、発声数の少ない音素について似ている音素を、音素の分類情報及び、話者クラスタリング情報を記憶する音素分類木及びクラスタ情報記憶部の分類に基づきグループ化し、複数の音素グループと同時に、前記音素分類木及びクラスタ情報記憶部から得たそれに対する話者クラスタリング情報を出力する。

【0025】ステップ2：ステップ1から受け取った音素グループ及び話者クラスタリング情報毎に、前記音素グループに対応する適応話者発声データを、複数話者の発声が予め記憶されている適応話者発声データ記憶部より読み出し、話者クラスタリング情報に対して必要に応じて学習話者発声データを前記学習話者発声データ記憶部から読み出し、クラスタごとにHMMを構成し、もっとも適当なクラスタの情報を出力する。

【0026】ステップ3：ステップ2で出力された音素グループ毎の学習話者発声データクラスタの情報から、必要に応じて学習データを前記学習話者発声データ記憶部から読み出し、適応元のHMMを、適応する元のHMMが記憶されている適応HMM記憶部より読み出し、適応を行ないその結果のHMMを出力する。

【0027】

【実施例】次に、本発明の実施例について図面を参照して詳細に説明する。図1は、本発明の第1の実施例の構成を示す図である。図1を参照すると、本発明の第1の実施例は、発声データ、標準パターンHMM（隠れマルコフモデル）などを保存する記憶装置1と、データ処理装置2とを備えて構成されている。

【0028】記憶装置1は、適応話者発声データ記憶部11と、音素分類木記憶部12と、学習話者発声データ記憶部13と、適応HMM記憶部14とを含む。

【0029】データ処理装置2は、音素グルーピング手段21と、話者クラスタリング手段22と、クラスタ選択手段23と、HMM適応手段24とを含む。

【0030】不図示の音声入力手段から入力されたデジタル信号に変換された、HMMを適応する対象の適応話者の発声データを、適応話者発声データ記憶部11に記憶される。

【0031】音素分類木記憶部12には全音素の分類情報が予め記憶されている。

【0032】学習話者発声データ記憶部13には、多数の話者の発声が予め記憶されている。適応HMM記憶部14には、適応する元のHMMが記憶される。

【0033】音素グルーピング手段21は、適応話者発声データを適応話者発声データ記憶部11より読み出し、発声数の少ない音素について、類似している音素を、音素分類木記憶部12の分類に基づきグループ化し、複数の音素グループを出力する。

【0034】話者クラスタリング手段22は、音素グルーピング手段21から出力された複数の音素グループのそれぞれについて、学習話者発声データを学習話者発声データ記憶部12より読み出して、話者単位で、クラスタリングを行ない、クラスタ化された学習話者発声データを、音素グループ毎に出力する。

【0035】クラスタ選択手段23は、音素グループ毎に、その音素グループに対応する適応話者発声データを、適応話者発声データ記憶部11より読み出して、入力された学習話者発声データのクラスタのそれぞれと比較し、最も適当なクラスタを出力する。

【0036】HMM適応手段24は、音素グループ毎の学習話者発声データクラスタを用いて、適応元のHMMを、適応HMM記憶部13より読み出し、適応を行ない、その結果のHMMを出力する。

【0037】次に、図1を参照して、本実施例の全体の動作について詳細に説明する。

【0038】音素グルーピング手段21で音素グループを作るに際して、音素分類木記憶部12に記憶された音素分類木を用いる。図1に、音素分類木の具体的な一例を示す。図3を参照すると、ルートに全音素があり、その子ノードに子音、母音があり、子音は、その子ノードとして有声音、無声音を持ち、母音は「A」、「I」、「U」、「E」、「O」をリーフとして持つ。以下、音素グルーピング手段21の動作を説明する。

【0039】まず、音素分類木の最上階層のノードに音素を分類し、それらのノード全てについて下記のステップ(a)以下の探索を開始する。図3に示す例では、最上階層のノードは一つであるため、そこに全音素が分類され、最初に唯一の探索ノードとして探索が開始される。

【0040】ステップa：各ノードにおいて、ステップb：子ノードがない場合、そのノードを分類結果の一つとし、そのノードの探索を終了する。

【0041】ステップc：子ノードがある場合、そのノードの子ノードそれぞれについて各音素分類木記憶部11から探し、発声数が十分あるかどうかを調べてゆく。

【0042】ここで、「十分ある」とは、個数あるいは個数とその記述長の線形和が、予め定められた閾値以上であることをいう。

【0043】その後、ステップd：一つの子ノードでも発声数が十分でない場合、その子ノードの親ノードを分類結果の一つとし、そのノードの探索を終了する。

【0044】ステップe：すべての子ノードで発声数が十分量ある場合、その各子ノードについて、ステップa以降の探索を再帰的に行なう。

【0045】ステップf：すべてのノードでの探索が終われば、出力を行ない動作を終了する。

【0046】以上の処理により、音素グルーピング手段21は、各音素分類結果に対する適応話者発声が少なくなりすぎない程度に細かく音素をグループ分けできる。

【0047】話者クラスタリング手段22は、ノード集合の個々に対し、そのノードに分類されている音素に対応する学習話者発声データを学習話者発声データ記憶部13よりすべて読み出し、それらの音素のHMMを作成し、話者クラスタリングを行い、その結果、音素グループ毎にクラスタに別れた学習話者発声データを出力する。

【0048】クラスタ選択手段23は、音素グループ毎に、クラスタ分けされた学習データについてHMMを作成し、適応発声に対する尤度のもっとも大きいものを選択し、その学習発声データを出力する。結果、すべての音素に対し学習データが揃う。

【0049】HMM適応手段24は、入力の学習発声を用いて、適応HMM記憶部14のHMMを話者適応する

いは特定話者学習し、話者適応後のHMMを出力する。

【0050】なお、適応HMM記憶部14を用いず、入力学習発声のみを用いてHMMを作成する場合にもよい。

【0051】次に本発明の第2の実施例について説明する。図2は、本発明の第2の実施例の構成を示す図である。図2を参照すると、本発明の第2の実施例は、前記第1の実施例と同様、発声データ、HMMなどを保存する記憶装置1とデータ処理装置2を備えて構成されており、記憶装置1は、適応話者発声データ記憶部11と、音素分類木と対応する話者クラス情報記憶部15と、学習話者発声データ記憶部13と、適応HMM記憶部14を含む。

【0052】データ処理装置2は、音素グルーピング手段21と、クラスタ選択手段22と、HMM適応手段23を含む。

【0053】前記第1の実施例と相違している点は、音素分類木及び対応する話者クラス情報記憶部15である。これは、全音素の分類情報の他に、分類それぞれに対して予め前記第1の実施例と同様な話者クラスターリングを行っておき、その結果の情報を付加した木が記録されている。

【0054】ここで、話者クラスターリング結果の情報とは、学習話者発声データをクラスに分割できる情報をさす。この場合、学習話者発声データ記憶部13に、同データを記憶させ、必要に応じて各部で読み込むものとす。

【0055】また、話者クラスターリング結果の情報を、分割された学習発声とすると、学習話者発声データ記憶部13は必要ない。

【0056】図4は、この例を示す図であり、話者クラスターリング結果の情報を、クラスに分割された学習発声によって作成されたHMMとした場合の構成を示している。図4を参照すると、音素分類木と、各音素分類について、話者クラスターリング結果の情報をもとにクラスに分割された学習発声によって作成されたHMMを記憶する記憶部16を備えている。図4に示す実施例においても、学習話者発声データが、それによって作成されたHMMを指すものとすれば、以下の説明がそのまま当てはまる。

【0057】適応話者発声データ記憶部11は、HMMを適応する対象の話者の発声データを記憶している。適応HMM記憶部14は適応する元のHMMが記憶されている。

【0058】データ処理装置2における処理の流れを説明する。

【0059】まず音素グルーピング手段21は、入力された適応話者発声データを適応話者発声データ記憶部11より得て、発声数の少ない音素について似ている音素を記憶部12の分類に基づきグループ化し、複数の音素

グループと同時に、音素分類及びクラスタ情報記憶部12から得たそれに対する話者クラスターリング情報を出力する。この時、必要に応じて学習話者発声データ記憶部13よりデータを受け取る。音素のグループ決定法は、前記第1の実施例と同様である。

【0060】クラスタ選択手段22は、受け取った音素グループ及び話者クラスターリング情報毎に、その音素グループに対応する適応話者発声データを適応話者発声データ記憶部11より読み出し、話者クラスターリング情報に対して必要に応じて学習話者発声データを学習話者発声データ記憶部13から読み出し、クラスタごとにHMMを構成し、前者を後者のそれぞれと比較し、もっとも適当なクラスタの情報を出力する。

【0061】クラスタと適応話者発声の比較、選択法は、前記第1の実施例と同様である。

【0062】HMM適応手段23は、入力の音素グループ毎学習話者発声データクラスタの情報から、必要に応じて学習データを学習話者発声データ記憶部13から読み出し、適応元のHMMを適応HMM記憶部14より読み出し、適応を行ないその結果のHMMを出力する。前記第1の実施例と同様に、適応HMM記憶部14を用いず読み込んだ学習話者発声データだけでHMMを作成することもできる。図4は、この場合の構成を示す。

【0063】なお、本発明の第2の実施例では、あらかじめ話者クラスターリングを行っておくので、話者適応を何度也行う場合、速度面向上が得られる。

【0064】

【発明の効果】以上説明したように、本発明によれば下記記載の効果を得る。

【0065】本発明の第1の効果は、適応話者の発声に応じて細かい音素グループ別に話者クラスターリングを行ない、クラスタを選択することで、その結果適応話者により近い学習話者発声データを厳密に集めることができる、ということである。

【0066】本発明の第2の効果は、より性能のよい話者適応を行うことができる、ということである。

【図面の簡単な説明】

【図1】本発明の第1の実施例の構成を示すブロック図である。

【図2】本発明の第2の実施例の構成を示すブロック図である。

【図3】本発明の第1の実施例を説明するための図であり、音素分類木の一例を示す図である。

【図4】本発明の第3の実施例の構成を示すブロック図である。

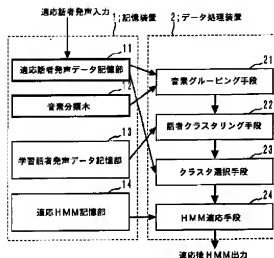
【符号の説明】

- 1 記憶装置
- 2 データ処理装置
- 11 適応話者発声データ記憶部
- 12 音素分類木記憶部

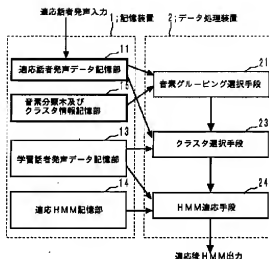
- 1 3 学習話者発声データ記憶部
1 4 適応HMM記憶部
1 5 音素分類木及びクラスタ情報記憶部
1 6 音素分類木及びクラスタ・HMM情報記憶部

- 2 1 音素グルーピング手段
2 2 話者クラスタリング手段
2 3 クラスタ選択手段
2 4 HMM適応手段

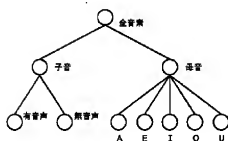
【図 1】



【図 2】



【図 3】



【図 4】

